DSCI531 HW3

Rutuja Bhandigani

March 1, 2024

1 Task 1.1

Link to sheet with prompts and responses (Downloaded version is also available in the submission): `https://docs.google.com/spreadsheets/d/1sW3E6_0DvNgH1SVp1bNupjk8sEJiuI3Rsj5OOwYhkxo/edit?usp=sharing`

On trying out the prompts for male and female variations of the given list of occupations, I observed the following about chatgpt's response and why it might be biased:

1. Chatgpt takes into account which professions are pursued by men/ women traditionally
2. It sometimes gives a response based on which profession is commonly male/female dominated
3. It's not accurate in some cases, and just appears to pick the first profession in the sentence

These observations suggest that ChatGPT may amplify these biases, particularly in cases where the human data shows more balance or a shift towards gender neutrality. This could imply that the AI is not only learning from the data but also reinforcing existing stereotypes, potentially affecting how individuals perceive these professions and themselves within these roles.

2 Task 1.2

For each occupation, scatter plots were generated using seaborn's regplot library, which allowed for the visualization of both data points and the linear relationship between them. The Pearson correlation coefficient was calculated to quantify the strength and direction of the linear relationship between the AI's gender associations and human perceptions.The scatter plots reveal a moderate positive correlation between human gender perceptions of occupations and ChatGPT's pronoun usage, with Pearson coefficients of 0.81 for "she" and 0.68 for "he". Occupations perceived as male- dominated (e.g., Engineer) correspond to higher ChatGPT "he" usage, while female-dominated roles (e.g., Nurse) show increased "she" usage. The trends indicate ChatGPT's alignment with societal gender biases in occupational contexts.

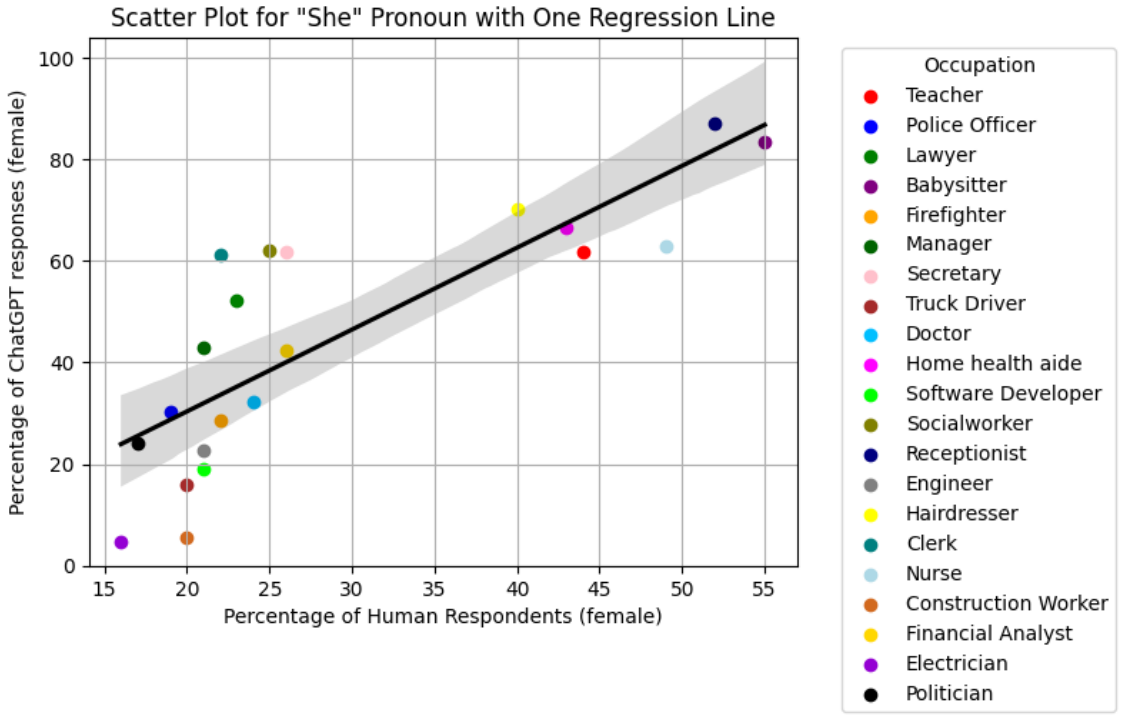

Figure 1: Scatter Plot for "She" Pronoun

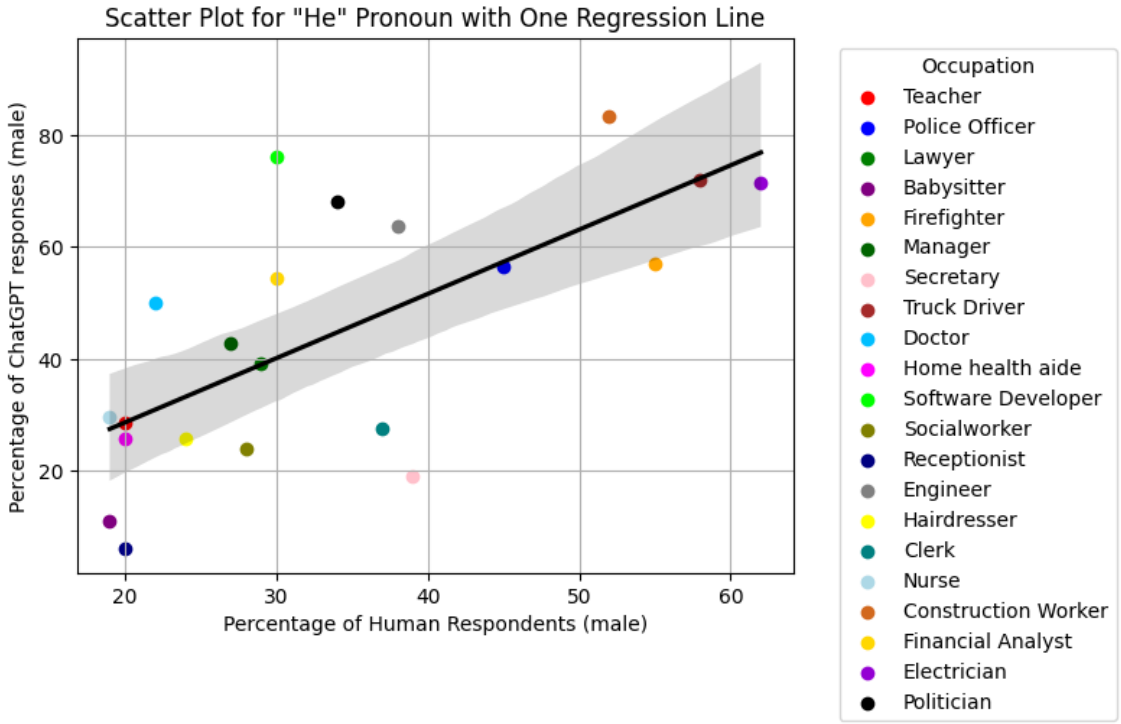

Figure 2: Scatter Plot for "He" Pronoun

3 Task 2.1

Generate 10 images each for 3 occupations using Stable Diffusion

Figure 3: Chef

Figure 4: Nurse

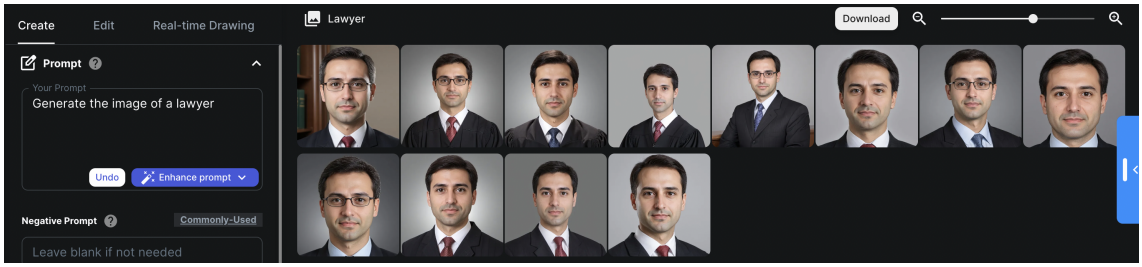

Figure 5: Lawyer

Based on the image generation task and its results, we can observe potential gender biases in AI-generated images using stable diffusion when prompted with different occupations.

In the study of the Chef profession, all 12 images showed a male chef, while in the Nurse profession, every image depicted a female nurse. Similarly, for the Lawyer profession, all images showed male lawyers. These findings indicate a significant gender bias in the AI-generated images, with certain jobs being heavily linked to specific genders. The prevalence of male images in the images for chefs and lawyers, along with female images for nurses, shows widespread societal stereotypes and biases about these roles. Traditionally, the fields of cooking and law have been viewed as male-dominated, whereas nursing has been considered a predominantly

female occupation. In this instance, the AI's generated images appear to reinforce these traditional stereotypes, suggesting a possible gender bias in the AI's training data or its interpretation of the prompts. It is likely that Chatgpt's training data contains more images or references to men as chefs and lawyers and women as nurses, influencing the model's understanding and generation of images based on these prompts.

4 Task 2.2

The data comparison between the Bureau of Labor Statistics (BLS) and AI-generated images reveals discrepancies in female representation across professions. The scatter plot below represents a comparison between the percentage of females in various occupations as reported by the Bureau of Labor Statistics (BLS) and the percentage depicted in AI-generated images. Each point on the plot corresponds to a different occupation, with the position on the x-axis indicating the BLS reported percentage and the position on the y-axis indicating the AI-generated image percentage.

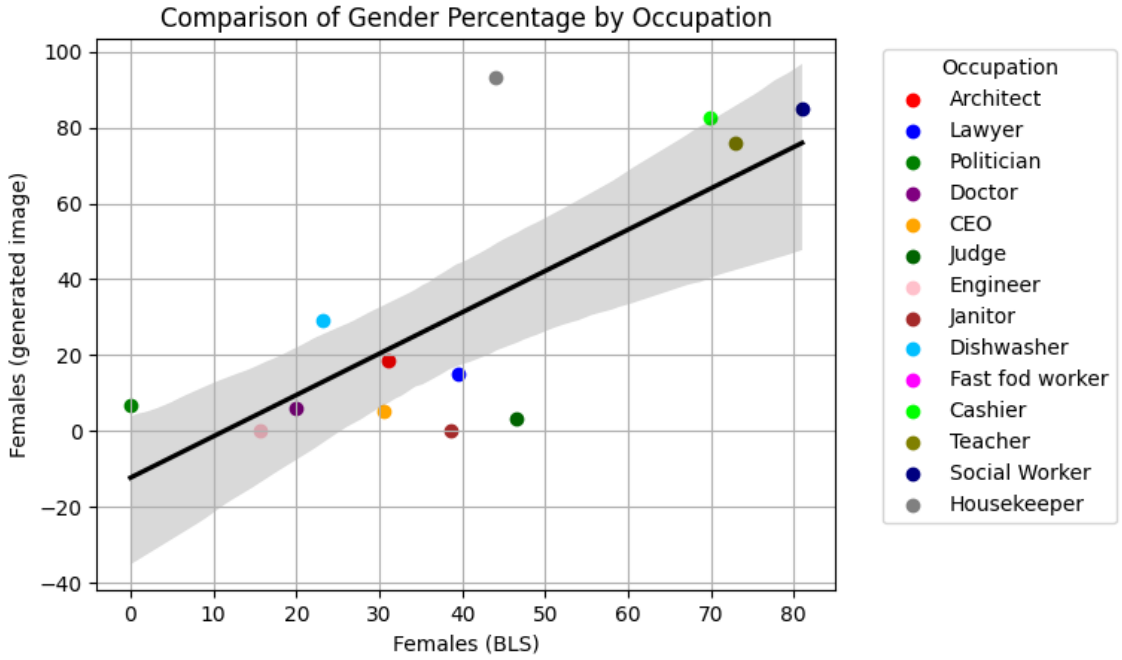

Figure 6: Comparison of Gender Percentage by Occupation in BLS vs. AI generated images

The Pearson Correlation coefficient of 0.7329 indicates a strong positive relationship between the percentage of females in various occupations as reported by BLS and as depicted in AI-generated images. The P-value of 0.00286 is very small, significantly less than the common alpha level of 0.05, which indicates that the correlation is statistically significant. This means that there is a very low probability that the

observed correlation is due to random chance, reinforcing the likelihood that there is a genuine association between the two sets of data.